



SUCAST

Ilustrações cedidas pelas autoras

Desvendando as vias de transdução de sinal da cana-de-açúcar

A cana-de-açúcar na era da genômica

Gláucia Mendes Souza

Prof^ª, Dr^ª do Departamento de Bioquímica, Instituto de Química, USP.
glmsouza@iq.usp.br

Aline Maria da Silva

Professora Associada do Departamento de Bioquímica, Instituto de Química, USP.
almsilva@iq.usp.br

Desde os tempos do Brasil colônia até os dias de hoje, a cultura de cana-de-açúcar tem sido uma grande fonte de riquezas para a economia brasileira. O Brasil é o maior produtor de cana-de-açúcar do mundo, com uma safra de 338 milhões de toneladas em 2001, o que equivale

a 27% da produção mundial¹. Aproximadamente, 60% da colheita destina-se à produção de álcool e o restante à produção de açúcar, mas existe a perspectiva de utilizar-se a cana-de-açúcar como biorreator na geração de energia elétrica, na produção de plásticos biodegradáveis, açúcares não calóricos e compostos químicos de interesse farmacêutico. A cana-de-açúcar é cultivada em 5 milhões de hectares distribuídos por todos os

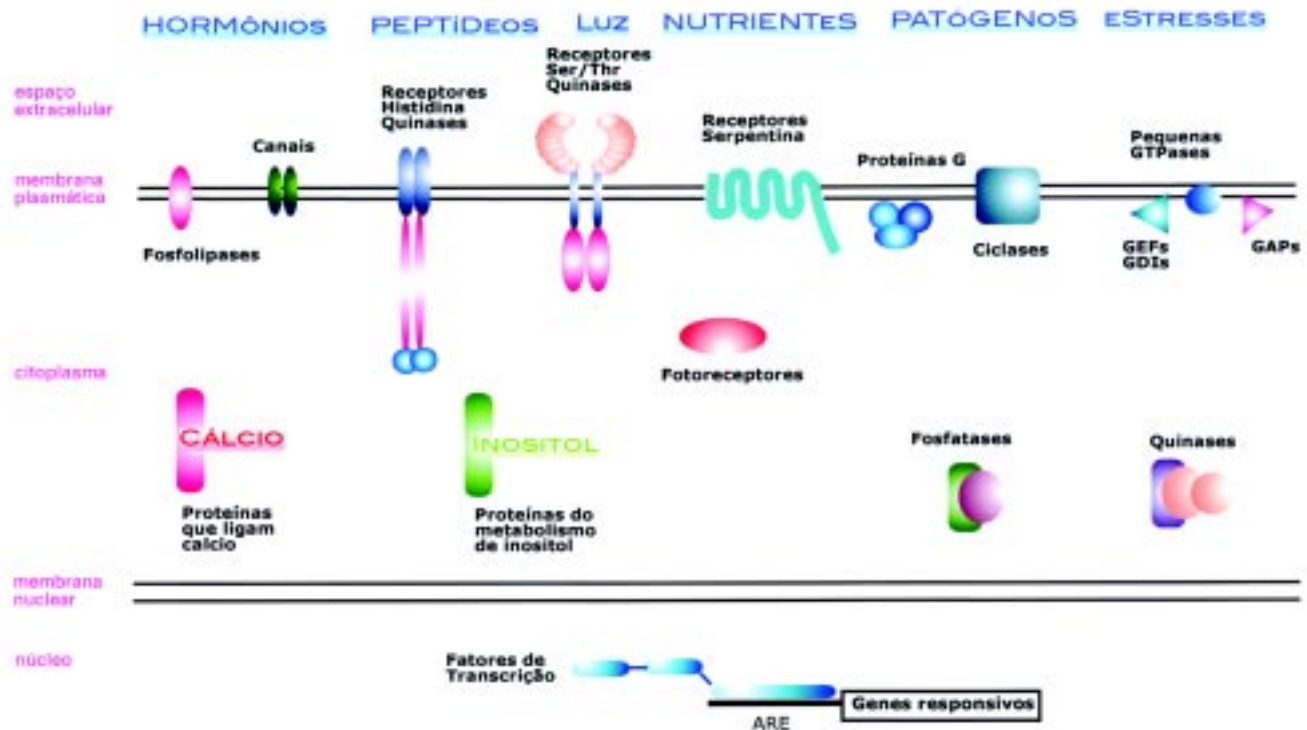


Figura 1 – Classes de proteínas de transdução de sinal catalogadas pelo SUCAST

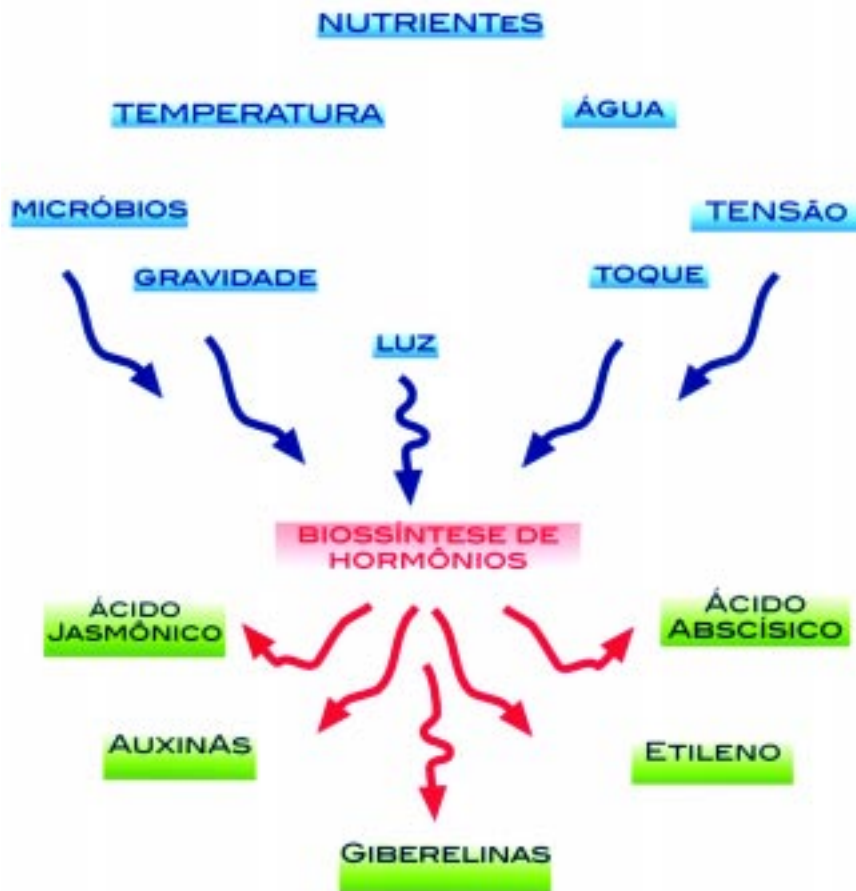


Figura 2 - Síntese de hormônios em resposta a fatores ambientais

estados brasileiros, mas é no estado de São Paulo que se concentra a maioria das lavouras, que representa metade da produção nacional. Alagoas, Pernambuco, Minas Gerais e Rio de Janeiro são também importantes produtores contribuindo juntos com quase 30% da safra brasileira.

O sucesso do cultivo da cana-de-açúcar se deve à utilização de variedades obtidas através de melhoramento genético clássico desenvolvido pelos centros de pesquisa e estações experimentais como o Centro de Ciências Agrárias da UFSCar, em Araras, o Instituto Agronômico de Campinas e a Copersucar. A cana-de-açúcar atualmente cultivada é originária de um cruzamento entre *Saccharum officinarum* e *Saccharum spontaneum*, que aumentou a produtividade e a resistência a doenças dos clones cultivados. O mapeamento genético dos cultivares mais utilizados está sendo realizado (Grivet e

Arruda, 2001) porém, a seleção de variedades mais produtivas, resistentes a pragas e doenças e adaptadas a ambientes diversos por inter cruzamentos é um processo demorado, que leva de 12 a 15 anos. Como podemos encurtar esse prazo? A genômica aponta o caminho. O seqüenciamento do genoma de várias plantas tem facilitado e acelerado a identificação de genes responsáveis por qualidades desejáveis, possibilitando a manipulação subsequente de genes de interesse através de técnicas de genética molecular. Na era da genômica as manipulações genéticas serão dirigidas, aumentando a eficiência de obtenção de variedades bem sucedidas. Essa nova revolução verde anuncia a obtenção de variedades resistentes a múltiplas doenças, mais adaptadas e produtivas o que deve diminuir dramaticamente as perdas na agricultura além de permitir o aproveitamento de solos até então não utilizáveis.

Em um futuro não muito distante, as culturas de cana-de-açúcar também serão beneficiadas por essa revolução, pois a cana-de-açúcar acaba de entrar na era genômica com um grande trunfo, representado pelas 250.000 ESTs geradas por 22 laboratórios que fazem parte do projeto SUCEST (<http://sucest.lad.ic.unicamp.br>; Fioravanti, 2000). ESTs (*Expressed Sequence Tags* ou Etiquetas de Seqüências Expressas) são seqüências de DNA que representam trechos de mRNAs (RNAs mensageiros que serão traduzidos em proteínas), revelando quais são os genes expressos em um tecido ou órgão, em uma dada situação fisiológica ou patológica. As ESTs da cana-de-açúcar não deixam por menos e representam milhares de genes expressos em seus diferentes órgãos, como raiz, colmo, folhas, flores e sementes, obtidos em vários estágios de desenvolvimento e submetidos a variações ambientais diversas incluindo interações com bactérias (Vettore et al., 2001). Essa informação preciosa está sendo cuidadosamente examinada por 48 laboratórios que se dedicam à mineração ou prospecção de dados, o chamado *data mining*. É dessa forma que os genes expressos revelados pelo projeto SUCEST estão sendo associados a uma função no crescimento, desenvolvimento, respostas a estresses e metabolismo da planta. A partir dessa análise, essas prováveis funções poderão ser associadas a características desejáveis na cana, que, eventualmente, possam ser manipuladas para a geração de novas variedades.

Como é feito o *data mining*

No *data mining* pesquisamos semelhanças entre genes da cana-de-açúcar e genes identificados em outras plantas ou em outros organismos, para designarmos uma provável função a esses genes, pois estima-se que 50% dos genes das plantas superiores tenham uma função semelhante à encontrada em outros organismos. A garimpagem é feita sobre as 250.000 ESTs da cana-de-

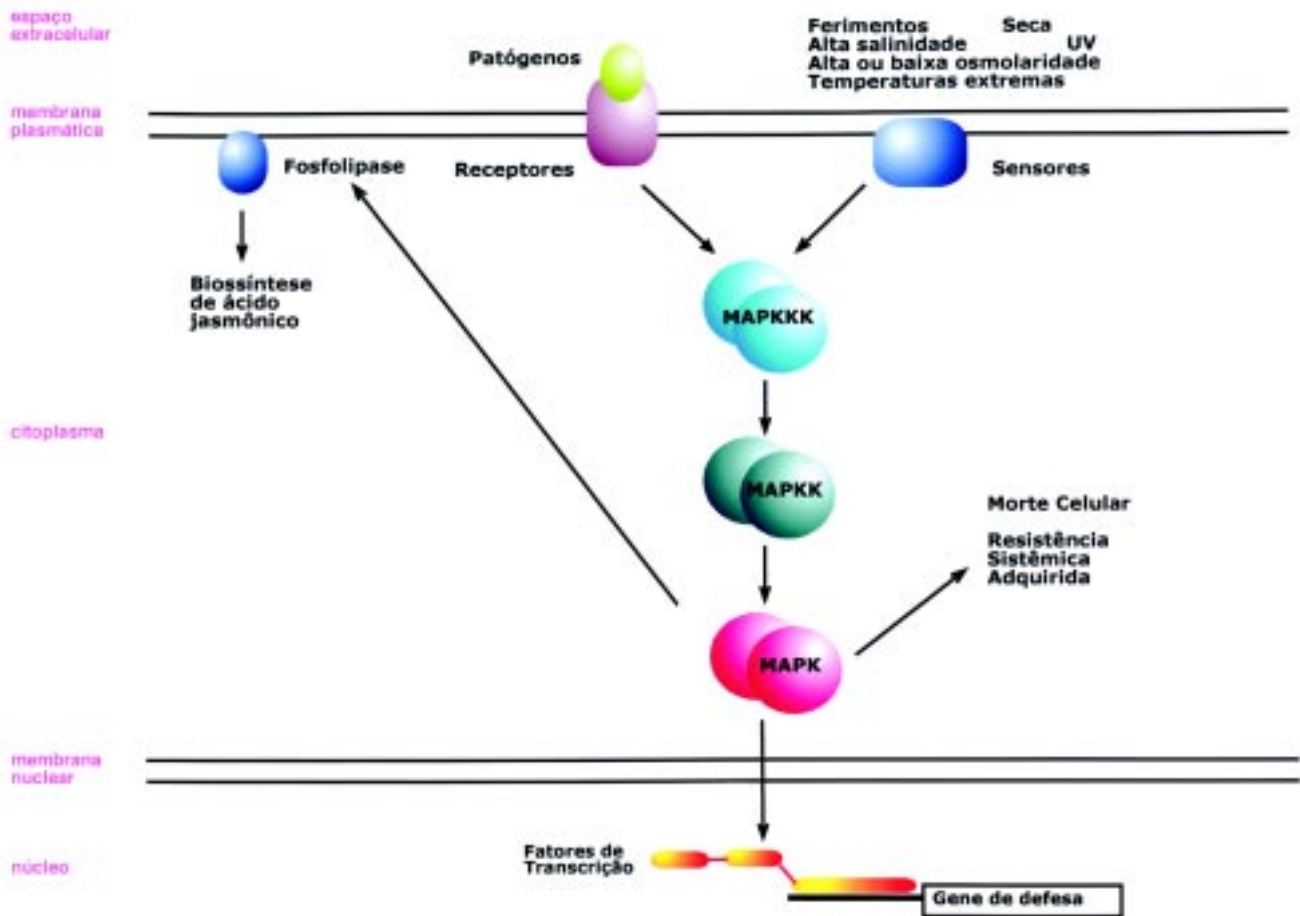


Figura 3 - Ativação da resposta de defesa através da cascata da MAP quinase (MAPK)

açúcar que foram previamente separadas, por semelhança entre suas seqüências, em 43.000 grupos de transcritos ou *clusters* (Pimentel & da Silva, 2001). Durante a análise também observamos quantas ESTs compõem o *cluster*, pois, em geral, o número de ESTs de um certo *cluster* indica, com boa confiança, o quanto aquele gene é expresso. Além disso, se um dado *cluster* é composto por ESTs provenientes de um tecido ou órgão em particular, temos uma boa indicação da prevalência de um certo transcrito em um tecido ou em determinada situação fisiológica ou patológica.

A tabela I relaciona as ferramentas de bioinformática utilizadas para inferência da função gênica a partir da análise de seqüências de genes e, naturalmente, das proteínas por eles codificadas. Invariavelmente, como primeiro passo, pesquisamos genes

de outros organismos que sejam similares as ESTs que estamos analisando. Alta similaridade com um gene de função conhecida indica uma boa probabilidade da EST corresponder a um gene de mesma função na cana-de-açúcar. A busca é realizada em bancos de dados públicos, em geral, com a ferramenta conhecida como BLAST. Essa busca pode ser automatizada e com o resultado obtido é possível gerar um banco de dados onde cada EST passa a ter uma provável função associada. Na prática, no entanto, a associação de uma função a um gene ou EST não é um processo trivial. No caso de famílias gênicas (genes parálogos), pequenas diferenças na seqüência das proteínas, que, codificadas por cada um dos parálogos, podem resultar em funções distintas e somente uma análise detalhada dessas seqüências conse-

gue distinguir membros de subcategorias. Nesses casos, além dos dados gerados pelo BLAST, também é inspecionado o alinhamento da seqüência de todos os membros da família, o que possibilita a indexação das ESTs ou dos *clusters* de ESTs a subfamílias e categorias gênicas específicas. Uma das ferramentas de alinhamentos mais utilizadas é o CLUSTAL. É desejável que domínios protéicos conservados sejam também pesquisados na seqüência de aminoácidos deduzida a partir da seqüência de nucleotídeos de cada um dos *clusters* de ESTs analisados. Nesse caso, as buscas são feitas em bancos de dados de domínios de proteínas como Pfam, PROSITE ou InterPro. Os resultados obtidos pelo conjunto dessas análises embasam a atribuição de uma provável função atribuída a um gene a partir apenas de dados do seqüenciamento.

Tabela I

Ferramentas de <i>data-mining</i>	Referência
BLAST - algoritmo que pesquisa similaridade entre seqüências	Altschul et al., 1997
CLUSTAL - algoritmo de alinhamento de seqüências	Jeanmougin et al., 1998
PROSITE - banco de dados de famílias e domínios	Hofmann et al., 1999
Pfam - banco de dados de famílias e domínios	Bateman et al., 1999
INTERPRO - banco de dados de famílias e domínios	Apweiler et al., 2001

SUCAST: Minerando a Transdução de Sinal da Cana-de-Açúcar

O projeto SUCAST (do inglês Sugar Cane Signal Transduction) tem como uma de suas propostas a identificação e catalogação de ESTs da cana-de-açúcar que estão envolvidos em vias de transdução de sinal. Durante o seu ciclo de vida, as plantas são constantemente bombardeadas por sinais ambientais e uma resposta adequada a cada um desses sinais é determinante para sua sobrevivência e para sua produtividade máxima. A compreensão das redes genéticas que orquestram tais respostas, que podem ser fisiológicas, bioquímicas, morfológicas ou de desenvolvimento é o foco de estudos recentes sobre transdução de sinal. Nesse contexto, a investigação dos mecanismos de integração de todos os sinais levou a identificação de vários hormônios que agem na planta de uma maneira tanto local quanto sistêmica, alterando o padrão de expressão de genes responsáveis por efetuar as mudanças necessárias em resposta aos sinais ambientais. Em muitos casos, o melhoramento genético de uma planta implica a manipulação de componentes que mediam a transdução de sinais, de modo que se explorem as redes de comunicação que detectam alterações do meio ambiente e que desencadeiam mudanças no padrão da expressão gênica.

A partir da análise de genomas completos que foram recentemente seqüenciados como o da levedura *Saccharomyces cerevisiae*, da mosca de fruta *Drosophila melanogaster*, do verme *Caenorhabditis elegans* e da planta *Arabidopsis thaliana* foi calculado que de 7% a 15% dos genes

de um organismo codificam para proteínas envolvidas na transdução de sinal. Se assumirmos que os 43.000 *clusters* ESTs da cana-de-açúcar representam, aproximadamente, o número de genes dessa planta, é razoável inferir que, pelo menos 4.000 de seus transcritos se encontrem na categoria de genes relacionados com a transdução de sinal (Souza et al., 2001). Utilizando-se as ferramentas delineadas na Tabela I, cerca de 900 *clusters* de ESTs já foram relacionados com diferentes vias de transdução de sinal da cana-de-açúcar e constam do catálogo do projeto SUCAST que está disponível no [website](http://sucest.lad.ic.unicamp.br/private/mining-reports/QG/QG-mining.htm) <http://sucest.lad.ic.unicamp.br/private/mining-reports/QG/QG-mining.htm>.

A figura 1 ilustra as diferentes classes de proteínas relacionadas com a transdução de sinal da cana-de-açúcar codificadas pelos *clusters* de ESTs que foram analisados até o momento. Essas classes estão relacionadas com o complexo sistema de sinalização que as plantas desenvolveram ao longo da sua evolução e que permite a sua adaptação a uma vasta gama de condições ambientais através, principalmente, da sinalização hormonal (para uma revisão sobre vias de transdução de sinal em plantas, ver Trewavas, 2000).

Sinais, receptores e mensageiros secundários da cana-de-açúcar

A ferramenta mais poderosa para identificar o papel dos hormônios na transdução de sinais foi a identificação de mutantes em suas vias biossintéticas. Através da utilização de técnicas de genética molecular e complementação de mutantes, muitas das enzimas de síntese e degradação de hormônios foram associadas a processos de cres-

cimento, envelhecimento, desenvolvimento, diferenciação, amadurecimento, dormência, resposta a ferimentos e defesa contra doenças. Entre os hormônios vegetais encontramos o gás etileno, as giberelinas, o ácido jasmônico, o ácido abscísico e as auxinas. Como esperado, a maioria das enzimas que participa na biossíntese desses hormônios está representada no conjunto de ESTs da cana-de-açúcar.

Acredita-se que a integração dos sinais ambientais durante as várias fases do ciclo de vida da planta se dê pela ação coordenada e simultânea de vários hormônios, uma vez que as mutações na via de síntese de um hormônio alteram, em muitos casos, a síntese de outro. A obtenção de plantas transgênicas alteradas nas vias de síntese de hormônios é de interesse econômico e tem sido alvo de intensa pesquisa. Como esquematizado na figura 2, os sinais ambientais podem ser ferimentos, peptídeos, produtos derivados de microorganismos e patógenos, além de outros agentes do meio ambiente como luz, gravidade, temperatura, vento, água, nutrientes e minerais do solo.

A maioria da recepção de sinais ocorre na da membrana celular, mas existem alguns exemplos de receptores intracelulares como os fotoreceptores que regulam o ritmo circadiano. Os receptores comumente encontrados são receptores ligados a proteínas G, ligados a canais de íons e os que possuem atividade enzimática. Essa última categoria é a mais freqüente na cana-de-açúcar, sendo principalmente representada pelos receptores com atividade de proteína quinase (atividade enzimática responsável pela fosforilação de certos resí-

duos de aminoácidos em substratos protéicos). O catálogo SUCAST relaciona *clusters* de ESTs de, pelo menos, 200 receptores, sendo, aproximadamente, 10% do tipo histidina quinase (fosforilam resíduos de histidina) e 50% do tipo serina/treonina quinase (fosforilam resíduos de serina ou treonina). Uma análise mais detalhada dos domínios das proteínas codificadas por essas ESTs revelou, nesse segundo grupo, uma subfamília numerosa que contém domínios ricos em leucina chamados LRR (do inglês, *Leucine Rich Repeat*). Essa classe de receptores constitui a base molecular do reconhecimento de patógenos pelos genes R de resistência a doenças onde o domínio LRR seria responsável pela mediação de interações entre proteínas. Domínios LRR também são encontrados em receptores do tipo CLAVATA1 e ERECTA, que regulam o crescimento do meristema.

As informações recebidas pelos receptores extracelulares são transmitidas para alvos intracelulares através dos chamados mensageiros secundários, que, em geral, são moléculas difusíveis. Mensageiros secundários da família do fosfatidilinositol estão presentes na cana, uma vez que as ESTs que codificam as enzimas da via de síntese dessas moléculas foram detectadas. Também identificamos ESTs relacionadas com o metabolismo do íon cálcio, como canais de cálcio, calmodulina, calreticulina e calnexina, sugerindo que esse importante mensageiro secundário é ativo na transmissão de sinais na cana-de-açúcar. Nas plantas, o cálcio é um sinal proeminente e as interações intracelulares reguladas por esse cátion são bastante complexas. Mudanças nos níveis de cálcio mediadas por canais de cálcio estão associadas com o início de algumas respostas, como o fechamento de estômatos, re-orientação do crescimento do tubo do pólen e aumento da espessura da parede em resposta ao vento.

Por outro lado, mensageiros secundários relevantes para outros organismos, como cAMP, cGMP e cADPr, parecem não estar presentes

em plantas e tampouco na cana-de-açúcar, pois não verificamos a presença de ESTs que codificam proteínas relacionadas ao metabolismo destes nucleotídeos entre o conjunto de ESTs seqüenciados no projeto SUCEST.

A síntese de mensageiros secundários após a ativação de certos tipos de receptores é mediada por um trio de subunidades protéicas (α , β e γ) conhecido como proteína G (porque se associa ao nucleotídeo GTP), e também por proteínas capazes de hidrolizar esse nucleotídeo, as chamadas GTPases pequenas. O processo de acoplamento e ativação dessas últimas é finamente regulado por proteínas acessórias, chamadas de GAPs, GEFs e GDIs, que funcionam como ativadores, dissociadores e inibidores da atividade GTPásica. Todos esses elementos estão codificados por ESTs da cana-de-açúcar e foram catalogados no projeto SUCAST, incluindo exemplos das três subunidades das proteínas G (α , β e γ) e todas as classes de uma enorme família de GTPases pequenas, além dos seus respectivos GAPs e GDIs

Proteínas quinases, fosfatases e fatores de transcrição

A grande maioria das vias de transdução de sinal engloba uma cascata de eventos de fosforilação e desfosforilação de proteínas catalisadas pelas proteínas quinases e proteínas fosfatases, respectivamente. A ativação dessas cascatas de fosforilação media respostas a estímulos distintos, como luz, agressão por patógenos, reguladores de crescimento, estresses variados e de privação de nutrientes, que são percebidos por receptores, como descrito anteriormente. Centenas de proteínas quinases e dezenas de proteínas fosfatases distintas estão representadas entre os *clusters* de ESTs da cana-de-açúcar, demonstrando que a fosforilação reversível de proteínas é um mecanismo de regulação importante também nessa planta. Entre as cascatas de fosforilação que ocorrem na cana-de-açúcar, destacamos a via da MAP quinase, que é uma das mais conhecidas em plantas e representa um exemplo clássico de cascata de fosforilação. O módulo básico da cascata é constituído

de três quinases chamadas MAPK, MAPKK (ou MEK) e MAPKKK (ou MEKK). Ao menos oito MAPKs foram identificadas entre os *clusters* de ESTs da cana-de-açúcar. Está postulado que sinais extracelulares, como os indicados na Figura 3, são captados por receptores, na sua maioria ainda desconhecidos, levando a ativação da MAPKKK, que, fosforila a MAPKK, que, por sua vez, fosforila a MAPK. Esta última proteína quinase provoca ativação da transcrição de genes de defesa e proteção a estresses, por exemplo. Além da ativação verificada diretamente sobre a atividade de proteína quinase dessas enzimas, foi verificado que os genes que as codificam tem a transcrição ativada em resposta a vários sinais e patógenos, conferindo resistência sistêmica a doenças.

A transdução do sinal extracelular culmina com mudanças no padrão de expressão de genes que codificam proteínas executoras da ação final que, em última instância, são as responsáveis pelas mudanças fisiológicas necessárias para os processos de adaptação, defesa, crescimento ou desenvolvimento em resposta aos distintos estímulos. A transcrição gênica é regulada por fatores de transcrição divididos em famílias, de acordo com os domínios protéicos característicos. Os mais comuns são os fatores de transcrição do tipo hélice-volta-hélice, com zíperes básicos de leucina, com dedos de zinco e com motivos HMG (do inglês, *high-mobility group*). Duzentos e cinquenta fatores de transcrição foram identificados até o momento entre os *clusters* de ESTs da cana-de-açúcar, incluindo um grande número de genes homeóticos que codificam fatores do tipo hélice-volta-hélice e atuam como mestres na regulação do desenvolvimento.

Como a genômica pode contribuir para o melhoramento da cana-de-açúcar?

A quantidade de informação gerada por um projeto da magnitude do SUCEST é imensa e projetos de prospecção de dados, como o SU-

CAST, são essenciais para que as informações relevantes para a pesquisa básica e para o melhoramento da cana-de-açúcar possam ser extraídas. O SUCAST tem como uma de suas metas a organização de todas as informações da estrutura e da função de genes e proteínas que compõem e regulam as vias de transdução de sinal da cana-de-açúcar, em um banco de dados que facilite futuras análises funcionais. Para essas análises, o projeto pretende incluir dados sobre a expressão dessa categoria de genes e que serão obtidos através da tecnologia de *microarrays* de DNA em lâminas de vidro (*chips* de DNA). Utilizando essa tecnologia, será possível investigar o perfil da expressão dos genes relacionados com a transdução de sinal em resposta a sinais ambientais variados, tais como estresses bióticos ou abióticos, além da comparação da expressão desse conjunto de genes entre variedades de cana-açúcar de interesse agrícola.

Além dos genes descritos neste artigo, outros projetos de *data mining*, que são parte do projeto SUCEST, já identificaram genes envolvidos na resistência a pragas, estresses diversos, metabolismo de açúcares, absorção de nutrientes, regulação do ciclo celular, assimilação de nitrogênio, tolerância ao alumínio e desenvolvimento da planta. A comparação do padrão de expressão desses genes com aqueles relacionados com a transdução de sinal, e que relatamos aqui, fornecerá pistas diretas sobre quais os módulos de sinalização é que regulam os processos fisiológicos e patológicos relevantes para a biologia da cana-de-açúcar. Ademais, aproximadamente, 46% das ESTs da cana-de-açúcar geradas pelo SUCEST não estão descritas em bases de dados públicas. Essas seqüências são inéditas e correspondem a genes ainda completamente desconhecidos! A engenharia genética da cana-de-açúcar e de outras gramíneas tem, portanto, um futuro garantido e promissor. As informações

derivadas da análise das 250.000 ESTs da cana-de-açúcar somadas aos dados de seqüenciamento de ESTs e genoma do arroz e do milho, que estão sendo gradualmente disponibilizados, compõem uma base sólida para a elucidação dos mecanismos de defesa, de adaptação a solos e de crescimento dessas plantas em regiões de cultivo diferenciadas.

Informações adicionais

Detalhes sobre o trabalho de *data mining* do projeto SUCAST podem ser encontrados em Souza et al., 2001 e nos seguintes endereços:

SUCAST: <http://sucest.lad.ic.unicamp.br/private/mining-reports/QG/QG-mining.htm>

SUCEST: <http://sucest.lad.ic.unicamp.br>

Informações sobre as autoras

Gláucia Mendes Souza é doutora em Bioquímica pela Universidade de São Paulo, fez pós-doutorado no La Jolla Cancer Research Foundation, em San Diego e no Baylor College of Medicine, em Houston, EUA. Atualmente é professora doutora no Departamento de Bioquímica do Instituto de Química da USP, São Paulo. E-mail: glmsouza@iq.usp.br

Aline Maria da Silva é doutora em Bioquímica pela Universidade de São Paulo, fez pós-doutorado na Saint Louis University, em Saint Louis, EUA. É livre-docente em Bioquímica pela USP e professora associada no Departamento de Bioquímica do Instituto de Química da USP, São Paulo. E-mail: almsilva@iq.usp.br

Referências

Altschul, S.F., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.

Apweiler, R., Attwood, T.K., Bairoch, A. et al. (2001). The InterPro database, an integrated documentati-

on resource for protein families, domains and functional sites. *Nucleic Acids Res.* 29:37-40.

Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Finn, R.D. & Sonnhammer E.L.L. (1999). The Pfam Protein Families Database. *Nucleic Acids Res.* 27:260-262.

Fioravanti, C. (2000). Os arquitetos da nova cana. *Revista Pesquisa FAPESP*, São Paulo, novembro 2000, pp 29-35.

Food and Agriculture Organization for the United Nations (FAO). www.apps.fao.org.

Grivet, L. & Arruda, P. (2001). Sugarcane genomics: depicting the complex genome of an important tropical crop. *Curr. Opin. Plant Biol.* 5: 122-127.

Hofmann, K., Bucher, P., Falquet, L., and Bairoch, A. (1999). The PROSITE database, its status in 1999. *Nucleic Acids Res.* 27:215-219.

Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends Biochem Sci.* 23: 403-405.

Ministério da Agricultura e do Abastecimento. www.agricultura.gov.br

Pimentel, G. & da Silva, F.R. (2001). Trimming and clustering sugarcane ESTs. *Genetics and Mol. Biol.* no prelo.

Souza, G. M., Simoes, A. C. Q., Oliveira, K. C., Garay, H. M., Fiorini, L. C., Gomes, F. S., Nishiyama-Junior, M. Y. & da Silva, A.M. (2001). SUCAST: prospecting signal transduction in sugarcane. *Genetics and Mol. Biol.* no prelo.

Trewavas, A. (2000). Signal Perception and Transduction. *Biochemistry and Molecular Biology of Plants*. Ed. Bob B. Buchanan, Wilhelm Gruissem e Russel L. Jones. American Society of Plant Physiologists.

Vettore, A., Kemper, E., da Silva, F. & Arruda, P. (2001). The libraries that made SUCEST *Genetics and Mol. Biol.* no prelo.

¹(Ministério da Agricultura e Abastecimento e Food and Agriculture Organization for the United Nations)